NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

Elliptic Curves Over Finite Fields

Sonia Balagopalan

Supervisor: Doctor Patrick C. McCarthy

Department of Mathematics

Faculty of Science

National University of Ireland, Maynooth

Maynooth, Ireland

October, 2009

Acknowledgments

I would like to thank my parents and sister for all their loving confidence and unwavering support. Without you I would not be.

I would like to thank Prof. N.S.N. Sastry, Prof. S. Inamdar, Prof G. Mishra, Prof. A. Sitaram, Prof. B. Sury, and all the faculty at ISI Bangalore for believing in me and teaching me what mathematics is. Without your encouragement and help, I would have had to give up my dream of becoming a mathematician a long time ago.

I would like to thank all my friends, especially Gabby, Kate, Billy, Brian, Ciaran, and Kurt for putting up with me and making Ireland a home to me. Your friendship means the world to me.

The mathematics department at NUI Maynooth have been invaluable in providing all kinds of support. I would like to thank Prof. Stephen Buckley and Dr. David Wraith for making everything possible. I want to thank Ms. Grainne O'Rourke and Dr. Ciaran Mac An Bhaird (again!), and everyone else in the department for being simply wonderful and coming through at all times, making this a great place to study and work.

It would not be enough to thank Dr. Pat McCarthy for being the perfect supervisor, most excellent teacher, and brilliant mathematician he is. It goes without saying that this thesis would not be possible without him. But what I am thankful for the most is for teaching me by example how to think about mathematics. If I ever become half as good a mathematician as you are, I will consider myself successful, and it will all be thanks to you.

Contents

Abstract

This thesis provides a self-contained introduction to elliptic curves accessible to advanced undergraduates and graduate students in mathematics, with emphasis on the the theory of elliptic curves over finite fields.

In Chapter 1, affine and projective planes are introduced. Chapter 2 introduces the theory of algebraic curves and the Weierstrass Normal Form of a cubic curve is derived. In Chapter 3 we define derivations on arbitrary polynomial rings, and prove the group law for elliptic curves. Chapter 4 discusses elliptic curves over finite fields and proves some results on counting points.

Chapter 1

Affine and Projective Planes

The real Euclidean plane, with its constituent points and lines, is a familiar geometric object. It is often envisioned as $\mathbb{R}^2$, the 2-dimensional vector space over the real numbers. Points can then be seen as elements (x, y) of $\mathbb{R}^2$, and lines as solutions to linear equations of the form $ax+by = c$. We have that any two points lie on a unique line, any two lines intersect in at most one point, and given a line and a point not on it, there exists a unique line through the latter which does not intersect the former. It would be interesting to study abstract structures that preserve these incidence properties of planes, while ignoring the additional algebraic structure vector spaces such as $\mathbb{R}^2$ have.

1.1 Affine Planes

A *point-line incidence structure* is a triple $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ consisting of two sets $\mathcal{P}$ and $\mathcal{L}$ and a relation $\mathcal{I}$ in $\mathcal{P} \times \mathcal{L}$. We call the elements of $\mathcal{P}$ points, and the elements of $\mathcal{L}$ lines. If $P \in \mathcal{P}$ and $L \in \mathcal{L}$ satisfy $P\mathcal{I}L$, we say that P

is incident with L, or that P lies on L.We shall now look at the conditions under which a point-line incidence structure forms an incidence geometry

Definition 1.1. An *incidence geometry* is a point-line incidence structure $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ satisfying the following axioms:

I1 Given any two points P_1, P_2, there exists a unique line L such that $P_1 \mathcal{I} L$ and $P_2 \mathcal{I} L$.

I2 Given any line $L \in \mathcal{L}$ there are at least two distinct points $P_1, P_2 \in \mathcal{P}$ such that $P_1 \mathcal{I} L$ and $P_2 \mathcal{I} L$.

A3 There exists three distinct points in $\mathcal{P}$ which are not all incident with the same line.

Note that axiom **I2** ensures that we cannot define lines incident with no point, or lines incident with one point. Such lines would not add any interesting structure to our geometry. Also note that axiom **I3** excludes the case where all lines are collinear

We call two lines L and L' in $\mathcal{L}$ parallel if they are either equal or disjoint (i.e, there are no points incident to both), and denote this by $L \parallel L'$. So by axiom **I1** above, if two lines L_1 and L_2 in $\mathcal{L}$ are not parallel, there exists a unique point $P \in \mathcal{P}$ such that $P \mathcal{I} L_1$ and $P \mathcal{I} L_2$.

Definition 1.2. An *affine plane* is an incidence geometry, or a point line incidence structure satisfying **I1**,**I2**,**I3**, also satisfying the following axiom:

A4 Given any point P and any line L_1, there exists a unique line L_2 such that $P \mathcal{I} L_2$, and L_1 and L_2 are parallel.

A4 is called the parallel postulate.

Example. Let K be a field, and K^2 be the set of all ordered pairs on K. Take $\mathcal{P}$ to be K^2. We define a line in K^2 to be the set of all solutions to an equation $aX + bY = c$ in K^2, where a and b are both not equal to 0. Let $\mathcal{L}$ to be the set of all lines in K^2, for all $P \in \mathcal{P}$ and $L \in \mathcal{L}$, define $P\mathcal{I}L$ if and only if $P \in L$.

Any two points are incident with a unique line. If (a_1, b_1) and (a_2, b_2) are two distinct points in K^2, there is a unique line incident with both these points given by the set of solutions to $(b_1 - b_2)(X - a_2) - (a_1 - a_2)(Y - b_2) = 0$.

Given any line and any point, there exists a second line incident with the given point and parallel to the first. Let L_1 be the line $\alpha X + \beta Y = \gamma$, and let P be the point (a, b). Now if $\gamma' = \alpha a + \beta b$, then (a, b) lies on the line L_2, given by $\alpha X + \beta Y = \gamma'$. L_1 and L_2 are parallel. Now assume L_3, given by the set of solutions to $\alpha' X + \beta' Y = \delta'$, with α', β' not both 0. It follows that (α, β) and (α, β) are linearly dependent over K. Therefore L3 has equation $\alpha X + \beta Y = \delta$, for some $\delta \in K$. Conversely all lines L_3 which are parallel to L_1 have this form. The set of all such lines, $\{\alpha X + \beta Y = \delta \mid \delta \in K\}$, partitions K^2. Therefore (a, b) will be incident with only line of this form, and L_2 is the unique line in $\mathcal{L}$ such that $P \in L_2$ and $L_1 \parallel L_2$.

Since K is a field, $0, 1 \in K$ are distinct elements. Therefore $(0, 0)$, $(1, 0)$ and $(0, 1)$ are three points in $\mathcal{P}$, which are not collinear, since $(0, 0)$ and $(1, 0)$ lie on the unique line $Y = 0$, and $(0, 1)$ does not lie on $Y = 0$. Hence there exist 3 non-collinear points.

So $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ is an affine plane, which we denote by $\mathbb{A}^2(K)$.

Of particular interest are affine planes of the form $\mathbb{A}^2(K)$, where K is a finite field, which have interesting number theoretic as well as combinatorial properties.

1.2 Projective Planes

Definition 1.3. A *projective plane* is a point-line incidence structure $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ which satises axioms **I1** and **I3** and additionally

P2 Given any line $L \in \mathcal{L}$ there are at least three distinct points which are incident with L.

P4 For any two lines $L_1, L_2 \in \mathcal{L}$, there is a point $P \in \mathcal{P}$ such that $P\mathcal{I}L_1$ and $P\mathcal{I}L_2$.

Example. Let $K^3 \setminus \{\mathbf{0}\}$ be the set of all "non-zero" ordered triples in K. Consider the equivalence relation $\sim$ on $K^3 \setminus \{\mathbf{0}\}$, given by $(x, y, z) \sim (\lambda x, \lambda y, \lambda z) \quad \forall \lambda \neq 0$, and let $\mathcal{P}'$ be the set of all equivalence classes in $\frac{K^3 \setminus \{\mathbf{0}\}}{\sim}$. We can denote the equivalence classes in $\frac{K^3 \setminus \{\mathbf{0}\}}{\sim}$ by homogeneous coordinates, with $\{\lambda(x, y, z) | \lambda \neq 0\}$ written as $[x : y : z]$. We define $\mathcal{L}'$ to be the set of all solutions to equations of the form $aX + bY - cZ = 0$, with all of $a, b, c \neq 0$. Note that since the above equation is homogeneous, it is well defined on $\frac{K^3 \setminus \{\mathbf{0}\}}{\sim}$. So we can define $\mathcal{I}'$ to be the obvious incidence relation. We can prove that $\mathbb{P}^2(K) = (\mathcal{P}', \mathcal{L}', \mathcal{I}')$ is a projective plane by using a similar argument to that in the proof that $\mathbb{A}^2(K)$ is an affine plane.

1.3 Projective Completion

Let $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ be an affine plane. Parallelism is an equivalence relation on the set of lines $\mathcal{L}$. Symmetry and reflexivity are straightforward. All we have to prove is transitivity. If any two or all three lines are equal, the equivalence follows from reflexivity. If $L_1, L_2, L_3 \in \mathcal{L}$ are distinct lines, such that $L_1 \cap L_2 = \phi$ and $L_2 \cap L_3 = \phi$, then if there exists $P \in L_1 \cap L_3$, there exists a *unique* line incident with P, parallel to L_2. But L_1 and L_3 are both incident with P and parallel to L_2, and $L_1 \neq L_3$. Therefore L_1 and L_3 are disjoint and hence parallel.

Let Ł$_\infty$ be the set of all equivalence classes in $\mathcal{L}$ under the equivalence relation $\|$. Let $\hat{L}$ denote the equivalence class of L in Ł$_\infty$. Define $(\widetilde{\mathcal{P}}, \widetilde{\mathcal{L}}, \widetilde{\mathcal{I}})$ as follows.
We dene the relation $\widetilde{\mathcal{I}}$ from $\widetilde{\mathcal{P}}$ to $\widetilde{\mathcal{I}}$ as follows:

1. For $P \in \mathcal{P}$ and $L \in \mathcal{L}$, $P\widetilde{\mathcal{I}}L \iff P\mathcal{I}L$.

2. For $Q \in$ Ł$_\infty$ and $L \in$ $(\mathcal{L})$, $Q\widetilde{\mathcal{I}}L \iff Q = \hat{L}$.

3. For $R \in \mathcal{P}$, $R\widetilde{\mathcal{I}}L_\infty \iff R \in L_\infty$.

L_∞ is called the line at infinity.

$(\widetilde{\mathcal{P}}, \widetilde{\mathcal{L}}, \widetilde{\mathcal{I}})$ is a projective plane. It is called the *projective completion* of $(\mathcal{P}, \mathcal{L}, \mathcal{I})$.

Any two distinct points in $\widetilde{\mathcal{P}}$ lie on a unique line. If $P_1, P_2 \in \mathcal{P}$, then they are incident with a unique line in $\mathcal{L}$, and no point in $\mathcal{P}$ is incident with L_∞. If $\hat{L}_1, \hat{L}_2 \in$ Ł$_\infty$, then they are both incident with L_∞, and if $L \in \mathcal{L}$, then L can only be incident with one equivalence class $\hat{L} \in$ Ł$_\infty$. If $P \in \mathcal{P}$ and

$\hat{L} \in Ł_\infty$, then there exists a line $L \in \mathcal{L}$ such that $P \in L \in \hat{L}$, and we have $P\widetilde{\mathcal{I}}L$ and $\hat{L}\widetilde{\mathcal{I}}L$.

Any two distinct lines in $\widetilde{\mathcal{L}}$ are coincident at a unique point. Let $L_1, L_2 \in \mathcal{L}$. If $L_1 \parallel L_2$, then $\hat{L}_1 = \hat{L}_2$, and we have $\hat{L}_1 \parallel L_1$ and $\hat{L}_1 \parallel L_2$, and there is no other common point of incidence. If $L_1 \nparallel L_2$, then by the axioms for the affine plane, there exists a unique point $P \in \mathcal{P}$ such that P is incident with L_1 and L_2, and we have $P\widetilde{\mathcal{I}}L_1$ and $P\widetilde{\mathcal{I}}L_2$. If $L \in \mathcal{L}$, then $\hat{L}$ is the unique point incident with L and L_∞.

Theorem 1.4. *The projective plane $\mathbb{P}^2(K) = (\mathcal{P}', \mathcal{L}', \mathcal{I}')$ is isomorphic to the projective completion of $\mathbb{A}^2(K)$.*

Proof. Recall that the point $[x : y : z]$ in $\mathbb{P}^2(K)$ represents the equivalence class $\{\lambda(x, y, z) | \lambda \in K \setminus 0\}$ of points in $W \setminus \mathbf{0}$. Thus we can choose a non-zero homogeneous coordinate of a point and fix it to be 1. Consider the map $\phi : K^2 \to \frac{K^3 \setminus \mathbf{0}}{\sim}$ defined by $\phi(x, y) = [x : y : 1]$. Recall that any line parallel to $aX + bY = c$ in $\mathcal{L}$ can be written as $aX + bY = c'$ for some $c' \in K$. Let $Ł_\infty$ denote the set of all parallel classes of lines in $\mathcal{L}$. So we can extend ϕ to $\widetilde{\mathcal{P}} = \mathcal{P} \cup Ł_\infty$ in the following manner. If L is given by the set of all solutions to $aX + bY = c$, we define $\phi(\hat{L}) = [a : b : 0]$. Now define $\psi : \mathcal{L} \to \mathcal{L}'$ such that if $L \in \mathcal{L}$ is given by the set of solutions to $aX + bY = c$ in K^2, and $L' \in \mathcal{L}'$ is given by the set of solutions to $aX + bY - cZ = 0$ in $\mathcal{P}'$, then $\psi : (L) \mapsto L'$. Note that both $a, b \neq 0$. So we can extend ψ to $\widetilde{\mathcal{L}} = (\mathcal{L}) \cup \{L_\infty\}$ by defining $\psi : L_\infty \mapsto L'_\infty$ where L'_∞ is the line given by all solutions to $z = 0$. Thus $\phi : \widetilde{\mathcal{P}} \to \mathcal{P}'$ and $\psi : \widetilde{\mathcal{L}} \to \mathcal{L}'$ are bijections. Also, whenever $P \in \widetilde{\mathcal{P}}$ and $L \in \widetilde{\mathcal{P}}$, such that $P\widetilde{\mathcal{I}}L$, then $\phi(P)\mathcal{I}'\psi(L)$. Thus the projective completion of

$\mathbb{A}^2(K)$ is isomorphic to $\mathbb{P}^2(K)$. $\square$

Chapter 2

Curves

Before we define curves, we establish two consequences of Hilbert's Nullstellensatz**??**.

Let K be any field. Consider the polynomial $f \in K[X_1, X_2, \ldots, X_n]$. Define $\mathcal{Z}_K(f)$ to be the set of all zeroes of f in K^n.

Proposition 2.1. *If f is not constant, and K is algebraically closed, then $\mathcal{Z}_K(f) \neq \phi$.*

Proof. This is a direct consequence of the Nullstellensatz. Since f is not constant, the ideal $\sqrt{(f)}$ is properly contained in $K[X_1, X_2, \ldots, X_n]$ and $\mathcal{Z}_K(f)$ is a non-empty subset of K^n. □

Note that K being algebraically closed is necessary in the above proposition. For example, the polynomial $X_1^2 + X_2^2 + 1$ has no zeroes in $\mathbb{R}^2$.

Proposition 2.2. *Let K be an algebraically closed field. Let $f, g \in K[X_1, X_2, \ldots, X_n]$ be nonconstant polynomials. Then $\mathcal{Z}_K(f) = \mathcal{Z}_K(g)$, if and only if f and g have the same irreducible factors.*

Proof. Let

$$f = \prod_{i=1}^{r} f_i^{\alpha_i} \text{ and } g = \prod_{j=1}^{s} g_j^{\beta_j}$$

be the factorisation into irreducibles of f and g respectively. Again, by the Nullstellensatz, $\sqrt{(f)} = \sqrt{(g)}$, i.e,

$$\prod_{i=1}^{r} f_i = \prod_{j=1}^{s} g_j.$$

Now since $K[X_1, X_2, \ldots, X_n]$ is a *unique factorisation domain*, irreducible elements are prime, and for all $1 \leq i \leq r$, we have $f_i = g_j$ for some $1 \leq s$. Similarly, each irreducible factor of g is an irreducible factor of f. Therefore f and g have the same irreducible factors. □

Again, the irreducibility of K is crucial. Consider the polynomials X^2+1 and $Y^2 + 1$ in $\mathbb{R}[X, Y]$. Both have zero set the empty set, but have no irreducible factors in common.

We shall now define an affine plane curve over K.

Definition 2.3. Let $\bar{K}$ be the algebraic closure of K. If $f \in K[X, Y]$ is a nonconstant polynomial such that f is irreducible in $\bar{K}[X, Y]$, then $\mathcal{Z}_K(f)$ is an affine plane curve.

If f is the product of two coprime polynomials, ie. $f = f_1 f_2$ for some $f_1, f_2 \in K[X, Y] \setminus K$ with $gcd(f_1, f_2)$ then $\mathcal{Z}(f) = \mathcal{Z}(f_1) \cup \mathcal{Z}(f_2)$, will not be a curve, as we do not want to include the unions of distinct curves in our definition of curves. When $f = (f_0)^n$ for some f_0 irreducible, $\mathcal{Z}(f) = \mathcal{Z}(f_0)$ and we can assume that the polynomial f was irreducible to start with.

It is well known that if f has degree 1 or 2, then $\mathcal{Z}(f)$ is either a line or a conic. The next simplest case is the cubic.

Before beginning to deal with the geometry of cubic curves, we shall look at plane curves as intersections of curves in projective planes with affine planes sitting in them.

2.1 Projective Curves

Consider the polynomial ring $K[X, Y, Z]$. $F(X, Y, Z) = 0$ takes solutions from K^3. We want the zero set of F to be well defined on $\mathbb{P}^2(K)$, we need $F(\lambda x, \lambda y, \lambda z) = 0$ for all $\lambda \in K \setminus 0$ whenever $F(x, y,) = 0$ and $(x, y, z) \in K^3 \setminus \mathbf{0}$. If F is a homogeneous polynomial, then $F(x, y, z) = 0 \Rightarrow F(\lambda x, \lambda y, \lambda z) = 0$.

Now,

$$\theta : \mathbb{A}^2(K) \hookrightarrow \mathbb{P}^2(K) \text{ via } \theta : (x, y) \mapsto x : y : 1 \text{ and}$$

$$\theta^{-1} : \mathbb{P}^2(K) \setminus \mathcal{Z}(Z = 0) \hookrightarrow \mathbb{A}^2(K) \text{ via } \theta^{-1} : [x : y : 1] \mapsto (x, y)$$

So if $F([x_0 : y_0 : z_0]) = 0$ and $z_0 \neq 0$, then $F([\frac{x_0}{z_0} : \frac{y_0}{z_0} : 1]) = 0$. Now if $F([x_0 : y_0 : z_0]) = 0$, then $F(x_0, y_0, z_0) = 0$ and $(\frac{x_0}{z_0}, \frac{y_0}{z_0}, 1)$ is a solution of $F(X, Y, Z) \in K[X, Y, Z]$. Suppose

$$F(X, Y, Z) = \sum_{i=0}^{d} \sum_{j=0}^{d-i} a_{ij} X^i Y^j Z^{d-i-j}$$

Letting $x = X/Z$ and $y = Y/Z$ we can define

$$f(x, y) = \frac{1}{Z^d} F(X, Y, Z) = \sum_{i=0}^{d} \sum_{j=0}^{d-i} a_{ij} x^i y^j \text{ and we have}$$

$$F(x_0, y_0, z_0) = 0, z_0 \neq 0 \Rightarrow f(\frac{x_0}{z_0}, \frac{y_0}{z_0}) = 0$$
$$Z(f) \supset \theta^{-1}(\mathcal{Z}(F) \setminus \mathcal{Z}(Z = 0)).$$

On the other hand, let $f(x, y) \in K[X, Y]$ be irreducible of degree d, and let $\mathcal{Z}(f)$ be the set of zeroes. Suppose

$$f(x, y) = \sum_{i=0}^{d} \sum_{j=0}^{d-i} \alpha_{ij} x^i y^j.$$

If we define $x = \frac{X}{Z}, y = \frac{Y}{Z}$, and

$$F(X, Y, Z) = Z^d f(x, y) = Z^d f(\frac{X}{Z}, \frac{Y}{Z}) = \sum_{i=0}^{d} \sum_{j=0}^{d-i} \alpha_{ij} X^i Y^j Z^{d-i-j},$$

we have $F(x_0, y_0, z_0) = 0$ whenever $f(\frac{x_0}{z_0}, \frac{y_0}{z_0}) = 0$. When $z_0 = 1$, we have $F([x_0 : y_0 : 1])$ whenever $f(x_0, y_0)$. Therefore, when $\alpha_{ij} = a_{ij}$

$$\mathcal{Z}(f) \subset \theta^{-1}(\mathcal{Z}(F) \setminus \mathcal{Z}(Z = 0)).$$

$$\therefore \mathcal{Z}(f) = \theta^{-1}(\mathcal{Z}(F) \setminus \mathcal{Z}(Z = 0))$$

Therefore,

$$\begin{aligned} \theta(\mathcal{Z}(f)) &= \mathcal{Z}(F) \setminus \mathcal{Z}(Z = 0) \\ &= \theta(\mathbb{A}^2(K)) \cap \mathcal{Z}(F) \end{aligned}$$

Now,

$$\begin{aligned} \mathcal{Z}(F) \cap \mathcal{Z}(Z = 0) &= \left\{ [x_0 : y_0 : 0] : \sum_{i+j=d} a_{ij} x_0^i y_0^j = 0, x_0 : y_0 \in \mathbb{P}^1(K) \right\} \\ &= \left\{ [1 : y_0 : 0] : \sum_{j=0}^{d} a_j y_0^j = 0 \right\} \cup [0 : 1 : 0] \end{aligned}$$

Therefore,

$$\mathcal{Z}(F) = \theta(\mathcal{Z}(F)) \cup \left\{ [1 : y_0 : 0] : \sum_{j=0}^{d} a_j y_0^j = 0 \right\} \cup [0 : 1 : 0] \tag{2.1}$$

We have a one-one correspondence between affine and projective plane curves. The map $f(x, y) \mapsto Z^{deg(f)} f(\frac{X}{Z}, \frac{Y}{Z})$ described above is called *homogenisation* and the inverse map is called *dehomogenisation*. We shall continue to make extensive use of these maps to study curves in whichever settings suit us.

2.2 The Weierstrass Normal Form

In this section, we show that every plane cubic curve on which we can find at least one point is birationally equivalent to the curve $f(x, y) = y^2 - (x^3 + ax + b) = 0$, when $char(K) \neq 2, 3$.

Let $f \in K[x, y]$, be an irreducible cubic polynomial. Let C be the corresponding plane curve in $\mathbb{A}^2(K)$, and $(x_0, y_0) \in C$. We can "shift" C such that $(0, 0)$ is a point on C, by the transformations $x \mapsto x + x_0$ and $y \mapsto y + y_0$. We shall call the corresponding irreducible cubic $f_1(x, y) = f(x + x_0, y + y_0)$. Denote

$$f_1(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3-i} a_{ij} x^i y^j.$$

Note that $a_{00} = f_1(0, 0) = 0$.

Now, we look at lines through $(0, 0)$ in $\mathbb{A}^2(K)$. All lines except the "vertical", (ie. $x = 0$), are given by zeroes of $l_t = y - tx$. This gives us another way of identifying the points on C, namely by (x, t) co-ordinates. For each $t \in K$, the points of intersection of $l_t = 0$ and $f_1 = 0$ are obtained by substituting $y = tx$. Since $(0, 0)$ is a point on the curve, the x coordinate of the

remaining (at most) two points will not be equal to 0. So we can divide by x and this gives $f_2 \in K[x,t]$, a quadratic in x. Defining

$$\begin{aligned} a(t) &= a_{30} + a_{21}t + a_{12}t^2 + a_{03}t^3 \\ b(t) &= a_{20} + a_{11}t + a_{02}t^2 \\ c(t) &= a_{10} + a_{01}t \end{aligned}$$

we have

$$f_2 = a(t)x^2 + b(t)x + c(t).$$

Since $char(K) \neq 2$, for all values of t such that $a(t) \neq 0$ we can multiply by $4a(t)$ and complete squares to solve $f_2 = 0$ for x, to get

$$(2a(t)x + b(t))^2 = b(t)^2 - 4a(t)c(t),$$

which gives at most two solutions for x for each value of t, corresponding to the points of intersection of $l_t = 0$ and $f_1 = 0$ other than $(0,0)$.

We can eliminate x from the above expression by substituting $s = 2a(t)x + b(t)$ to get $f_3 \in K[t,s]$, where

$$f_3(t,s) = s^2 - (b(t)^2 - 4a(t)c(t)).$$

Note that given t, for each of the possible values of x, we have a corresponding s. All we are actually doing here is choosing a new system of coordinates which is less messy than the previous one.

Define $p(t) = b(t)^2 - 4a(t)c(t)$. If the coefficient of t^4 in $p(t)$ is 0, then $p(t)$ is at most a cubic. If not, if $p(t_0) = 0$, we can write $p(t) = q(t)(t - t_0)$. Now for $t \neq t_0$, we can let $u = \frac{1}{(t-t_0)}$ and $v = \frac{s}{(t-t_0)^2}$. If $q(t) = r(t - t_0)$,

then $\frac{q(t)}{(t-t_0)^3} = \frac{r(t-t_0)}{(t-t_0)^3} = g(\frac{1}{(t-t_0)}) = g(u)$ is at most a cubic in u. Dividing f_3 by $(t-t_0)^4$ we have $f_4 \in K[u,v]$ such that

$$f_4(u,v) = v^2 - g(u).$$

Let g_2 be the coefficient of u^2 in g, since $char(K) \neq 3$, we can send u to $u - \frac{g_2}{3}$ to get $f_5(u,v) = f_4(u - \frac{g_2}{3}, v)$, $f_5 = 0$ giving

$$v^2 = h(u)$$

$h(u)$ a cubic with second degree coefficient 0. The set $\mathcal{Z}(f_5)$ of zeroes of f_5 is birationally equivalent to C.

We shall henceforth assume that all cubic polynomials $K[x,y]$, with $char(K) \neq 2, 3$, with at least one zero in $\mathbb{A}^2(K)$ are given by their Weierstrass normal form:

$$f(x,y) = x^3 + ax + b - y^2 \tag{2.2}$$

The above corresponds to the homogeneous polynomial in $K[X,Y,Z]$ given by:

$$F(X,Y,Z) = X^3 + aXZ^2 + bZ^3 - Y^2Z \tag{2.3}$$

Chapter 3

Elliptic Curves

3.1 Derivations

We give a formal algebraic definition of polynomial differentiation. Let R be a commutative ring with identity.

Definition 3.1. A map, or operator $D : R \longrightarrow R$ is called a *derivation* or *derivative* if

$$\begin{aligned} D(u+v) &= Du + Dv \text{ and} \\ D(uv) &= uDv + vDu \end{aligned}$$

for all $u, v \in R$. We call (R, D) a differential ring.

Let (R, D) be a differential ring. Note that $Du = Du + D0$ for all $u \in R$ and hence, $D0 = 0$ and $D1 = 1D1 + 1D1$ giving $D1 = 0$.

Note that it is possible to define more than one derivation on a ring. In particular, all R-linear combinations of derivations on R are themselves

derivations on R, ie., if D_1, D_2 are derivations on R and $u_1, u_2 \in R$, then $u_1D_1 + u_2D_2$ is a derivation.

We are particularly interested in the following maps on polynomial rings.

Proposition 3.2. *Let $R[X]$ be a polynomial ring. Then $D_X : R[X] \longrightarrow R[X]$, such that*

$$D_X : \sum_{i=0}^{d} a_i X^i \mapsto \sum_{i=1}^{d} i a_i X^{i-1}$$

is a derivation on $R[X]$.

Proof. Let

$$u = \sum_{i=0}^{m} a_i X^i \text{ and}$$

$$v = \sum_{i=0}^{n} b_i X^i.$$

Then

$$D(u+v) = \sum_{i=1}^{max\{m,n\}} i(a_i + b_i)X^{i-1} = Du + Dv$$

and we expand $D(uv)$ as

$$\begin{aligned} D(uv) &= D \sum_{j=0}^{m+n} \left(\sum_{i=0}^{j} a_i b_{j-i} \right) X^j \\ &= \sum_{j=1}^{m+n} \left(\sum_{i=0}^{j} j a_i b_{j-i} \right) X^{j-1} \end{aligned}$$

(Changing the index of summation j to $j-1$)

$$\begin{aligned} &= \sum_{j=0}^{m+n-1} \left(\sum_{i=0}^{j+1} (j+1) a_i b_{j-i+1} \right) X^j \\ &= \sum_{j=0}^{m+n-1} \sum_{i=0}^{j+1} \left(a_i (j-i+1) b_{j-i+1} + i a_i b_{j-i+1} \right) X^j \end{aligned}$$

Splitting the sum and rearranging, we get

$$\sum_{j=0}^{m+n-1}\left(\sum_{i=0}^{j+1} a_i X^i (j-i+1) b_{j-i+1} X^{j-i} + \sum_{i=0}^{j+1} i a_i X^{i-1} b_{j-i+1} X^{j-i+1}\right)$$

$$\begin{aligned}
&= \sum_{j=0}^{m+n-1}\left(\sum_{i=0}^{j} a_{j-i+1} X^{j-i+1} i b_i X^i + \sum_{i=0}^{j+1} i a_i X^{i-1} b_{j-i+1} X^{j-i+1}\right)\\
&= \left(\sum_{i=0}^{n} i b_i X^{i-1}\right)\left(\sum_{i=0}^{m} a_i X^i\right) + \left(\sum_{i=0}^{m} i a_i X^{i-1}\right)\left(\sum_{i=0}^{n} b_i X^i\right)\\
&= vDu + uDv
\end{aligned}$$

□

Note that in the previous case $Da = 0$ for all $a \in R$. We now describe partial differentiation.

Proposition 3.3. *Let (R, D) be a differential ring. Then the map $\Delta(D)$ or $\Delta : R[X] \longrightarrow R[X]$, such that*

$$\Delta(D) : \sum_{i=0}^{d} a_i X^i \mapsto \sum_{i=0}^{d} D a_i X^i$$

is a derivation on $R[X]$.

Proof. Let

$$u = \sum_{i=0}^{m} a_i X^i \text{ and}$$

$$v = \sum_{i=0}^{n} b_i X^i.$$

Then

$$\Delta(u+v) = \sum_{i=1}^{max\{m,n\}} D(a_i + b_i) X^i = \Delta u + \Delta v$$

and

$$\begin{aligned}
\Delta(uv) &= \Delta \sum_{j=0}^{m+n} \left(\sum_{i=0}^{j} a_i b_{j-i} \right) X^j \\
&= \sum_{j=0}^{m+n} \sum_{i=0}^{j} D(a_i b_{j-i}) X^j \\
&= \sum_{j=0}^{m+n} \sum_{i=0}^{j} \left(a_i D b_{j-i} + b_{j-i} D a_i \right) X^j \\
&= \sum_{j=0}^{m+n} \sum_{i=0}^{j} \left(a_i X^i D b_{j-i} X^{j-i} + b_{j-i} X^{j-i} D a_i X^i \right) \\
&= \left(\sum_{i=0}^{m} a_i X^i \right) \left(\sum_{i=0}^{n} D b_i X^i \right) + \left(\sum_{i=0}^{n} b_i X^i \right) \left(\sum_{i=0}^{m} D a_i X^i \right) \\
&= u \Delta v + v \Delta u
\end{aligned}$$

□

In the above case we have $\Delta X = 0$ and $\Delta a = Da$ for all $a \in R$. We can generalise as

Corollary. If (R, D) is a differential ring, $R[X_1, X_2]$ is the polynomial ring in X_1, X_2 over R, and $\Delta^1 : R[X_1] \longrightarrow R[X_1]$ is the map in Proposition 3.3, then $\Delta^{1,2} : R[X_1, X_2] \longrightarrow R[X_1, X_2]$ such that

$$\Delta^{\{1,2\}} : \sum_{i=0}^{d} a_i(X_1) X_2^i \mapsto \sum_{i=0}^{d} \Delta^1(a_i(X_1)) X_2^i$$

is a derivation. We can inductively define derivations Δ^S for any finite set S of variables.

We can now define partial derivatives of polynomials in more than one variable.

Definition 3.4. Let K be a field and let $K[X_1, \ldots, X_n]$ be the ring of polynomials in variables $X_1, \ldots, X_n$ over K. Let

$$D_i : K[X_i] \longrightarrow K[X_i]$$

be as in Proposition 3.2, $S_i = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$ and let

$$\Delta^{S_i} : K[X_i][S_i] \longrightarrow K[X_i][S_i]$$

be the map $\Delta^{S_i}(D_i)$ defined in Propositions 3.2 and 3.3. We define the *partial derivative with respect to* X_i of $f \in K[X_1, \ldots, X_n]$ by $\Delta^{S_i} f$.

The chain rule and rules of implicit differentiation analogous to the real variables case hold for differential rings.

3.2 Tangents, Multiplicities and the Weierstrass Normal Form

Now that we have defined the derivative of a polynomial in $K[x, y]$, we can obtain the Weierstrass normal form from a cubic in a rather more satisfying manner than the earlier one, given just one condition on f. We shall not prove all statements we make in this section, just the ones we need. Let $f(x, y) \in K[x, y]$, and $C = \mathcal{Z}(f) \in \mathbb{A}^2(K)$.

Let $P = (x_0, y_0) \in C$. P is a *singular point* of f if $\frac{df}{dx}(x_0, y_0) = 0$ and $\frac{df}{dy}(x_0, y_0) = 0$. A point not satisfying the above conditions is called a *nonsingular point*.

Let $P = (x_0, y_0)$ be a nonsingular point on C. Let $\alpha = \Delta_x f(x_0, y_0)$ and $\beta = \Delta_y f(x_0, y_0)$. Since P is nonsingular, the set l of points (x, y) such that

$$\alpha(x - x_0) + \beta(y - y_0) = 0 \tag{3.1}$$

is a line, and $(x_0, y_0) \in l$. We call l the *tangent line* to E at (x_0, y_0).

Note that this definition of the tangent line coincides with our geometric notion of tangent to a curve wherever a geometric picture makes sense.

If C, C' are two curves in $\mathbb{A}^2(K)$, and $C \cap C' \not\supseteq D$ for all curves D in $\mathbb{A}^2(K)$ which have more than one point, we say that C and C' do not have a common component. Suppose C and C' do not have a common component, and $P \in C \cap C'$ is a nonsingular point. Let

$$u, v : K \longrightarrow K$$

be polynomial maps such that $C \cap C' \subset \{(u(t), v(t)) | t \in K\}$ and $(u(0), v(0)) = (x_0, y_0) = P$.

parametrisation, definition of multiplicity, intersection multiplicity, inflection points, weierstrass form given that a cubic has at least one inflection point

3.3 Definition of an Elliptic Curve

Let K be a field, $Char(K) \neq 2, 3$ $f \in K[x, y]$ be an irreducible polynomial, and let $C = \mathcal{Z}(f) \subset \mathbb{A}^2(K)$ be the curve given by the zeroes of f. Assume that we are given a point on C, ie., we are given $(x_0, y_0) \in \mathbb{A}^2(K)$ with $f(x_0, y_0) = 0$.

If all points of C are nonsingular, then C is a *smooth* or *nonsingular curve* and we also call the associated polynomial nonsingular.

We are now in a position to define elliptic curves.

Definition 3.5. An elliptic curve over K is a set $E = \mathcal{Z}(F) \subset P^2(K)$ where F is a nonsingular homogeneous cubic in $K[X, Y, Z]$ given by the Weierstrass form of (2.3)

An alternative characterisation of an elliptic curve E is as the set $\mathcal{Z}(f) \cup \mathcal{O}$ where $f \in K[x, y]$ is a nonsingular cubic given by the Weierstrass normal form in (2.2) and $\mathcal{O}$ is "a point at infinity".

If f is a polynomial in Weierstrass form, ie.

$$f(x, y) = x^3 + ax + b - y^2,$$

then

$$\Delta_x f(x, y) = 3x^2 + a \text{ and } \Delta_y f(x, y) = -2y.$$

For E to be nonsingular, for all $(x_1, y_1) \in E$, either $3x_1^2 + a \neq 0$ or $y \neq 0$. In other words, if $3x_1^2 + a = 0$, then $x_1^3 + ax_1 + b$ can not be 0. Substituting $x_1^2 = \frac{-a}{3}$ we get $\frac{-2a}{3}x_1 \neq b$, which we square to get the equivalent condition $4a^3 + 27b^2 \neq 0$.

Recall that $4a^3+27b^2 \neq 0$ is precisely the condition that $f(x, 0) = x^3 + ax + b$ does not have a double root, ie., $f(x, 0)$ and $\Delta_x f(x, 0)$ have no common zeroes.

3.4 Intersections

Let E be the elliptic curve given by (2.2), and let $(x_1, y_1) \in E$. When $F \in K[X, Y, Z]$ is as in (2.3), $E \hookrightarrow \mathbb{P}^2(K)$ and $[x_1 : y_1 : 1] \in E$

Lemma 3.6. *If L is a line in $\mathbb{P}^2(K)$, then $\#(E \cap L) \leq 3$.*

Proof. Let

$$L = \alpha X + \beta Y + \gamma Z$$

We want all common solutions in $\mathbb{P}^2(K)$ of

$$\begin{aligned} X^3 + aXZ^2 + bZ^3 - Y^2Z &= 0 \text{ and} \\ \alpha X + \beta Y + \gamma Z &= 0 \end{aligned}$$

Substituting $Z = 0$ in the first equation gives $X = 0$, therefore $Y = 1$, and $[0 : 1 : 0]$ satisfies the second equation iff $\beta = 0$

Substituting $Z = 1$ in both equations gives

$$\begin{aligned} X^3 + aX + b - Y^2 &= 0 \text{ and} \\ \alpha X + \beta Y + \gamma &= 0 \end{aligned}$$

If $\beta \neq 0$, ie. $[0 : 1 : 0] \notin L$ then we can substitute $\frac{-\gamma - \alpha X}{\beta}$ for Y in the first equation to get a cubic in X which has at most 3 solutions.

If $\beta = 0$, ie. $[0 : 1 : 0] \in L$, $\alpha \neq 0$ and we have $X = \frac{-\gamma}{\alpha}$, which we substitute to get a quadratic in Y which has at most 2 solutions. □

In the rest of this section, we shall look at how different kinds of lines intersect E. We parametrise the line l through (x_1, y_1) and (x_2, y_2) by t. All

points on l are given by

$$l = \{(x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1)) | t \in K\} \tag{3.2}$$

The points in $l \cap E$ are given by substituting $(x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1))$ into $f(x, y) = 0$ and expanding, which gives

$$(x_1 + t(x_2 - x_1))^3 + a(x_1 + t(x_2 - x_1)) + b - (y_1 + t(y_2 - y_1))^2 = 0 \tag{3.3}$$

and solving for t.

3.4.1 Tangent lines

Let $P = (x_1, y_1) \in E$. Since E is nonsingular, we have a tangent line at P given by (3.1), whose coefficients are $\alpha_1 = \Delta_x f(x_1, y_1) = 3x_1^2 + a$ and $\beta_1 = \Delta_y f(x_1, y_1) = -2y_1$.

Now if $(x_2, y_2) \neq (x_1, y_1)$ is another point of l, then all solutions of (3.1) are given by (3.2) which we substitute into (2.2) to get (3.3).

We can determine the points of intersection of E and l by expanding (3.3) and solving the above cubic equation for t. Then the 0 degree term of (3.3) is

$$x_1^3 + ax_1 + b - y_1^2 = 0$$

since $(x_1, y_1) \in E$. The coefficient of t in (3.3) is

$$(3x_1^2 + a)(x_2 - x_1) - 2y_1(y_2 - y_1) = \alpha_1(x_2 - x_1) + \beta_1(y_2 - y_1) = 0$$

since $(x_2, y_2) \in l$. Therefore, $t = 0$ is a double root of (3.3) and if $x_2 \neq x_1$, the third root is

$$t = \frac{(y_2 - y_1)^2 - 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}$$

The value of t depends on the choice of (x_2, y_2). But on substituting for t in $(x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1))$, we see that the third point of intersection (x_0, y_0) is given by

$$\begin{aligned} x_0 &= x_1 + \frac{(y_2 - y_1)^2 - 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}(x_2 - x_1) \\ &= \frac{(y_2 - y_1)^2}{(x_2 - x_1)^2} - 2x_1 \\ y_0 &= y_1 + \frac{(y_2 - y_1)^2 - 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}(y_2 - y_1) \\ &= \frac{(y_2 - y_1)^3}{(x_2 - x_1)^3} - 3x_1\frac{(y_2 - y_1)}{(x_2 - x_1)} + y_1 \end{aligned} \tag{3.4}$$

$\frac{y_2 - y_1}{x_2 - x_1}$ is the slope of the tangent line l, which is independent of the choice of (x_2, y_2).

Also, the case $x_2 = x_1$ occurs if and only if the tangent line at (x_1, y_1) is "vertical", or l is the line $x - x_1 = 0$. This in turn occurs iff $y_1 = 0$, or x_1 is a root of the polynomial $f(x, 0) = x^3 + ax + b$. In this case, l and $\mathbb{A}^2(K)$ intersect only at (x_1, y_1). But $y_1 = 0 \Rightarrow \beta_1 = 0$. So if L is the projective closure of l in $\mathbb{P}^2(K)$, then $[0 : 1 : 0] \in L$. Therefore, $[0 : 1 : 0] \in L \cap E \subset \mathbb{P}^2(K)$.

Coming back to our earlier observation that $t = 0$ is a double root of (3.3). $t = 0$ in the parametrisation of l above corresponds to the point $(x_1, y_1) \in E \cap l$. The natural interpretation of this is that the line l meets the curve E at (x_1, y_1) at least twice. We say that the *intersection multiplicity* of l with E at (x_1, y_1) is at least 2. Also note that the tangent line we defined is the unique line through (x_1, y_1) such that $t = 0$ is a multiple root. This coincides with the intuitive idea of the tangent as the best linear approximation to a curve. Also, recall that we insisted on nonsingularity of

E. The uniqueness of the tangent line at every point is guaranteed by the fact that E is nonsingular.

3.4.2 Secant lines

Now, let (x_1, y_1) and (x_2, y_2) be distinct points on $E \setminus \mathcal{O}$. Again, the line l joining the two points can be given by (3.2).

Substituting into (2.2), we have (3.3) which we expand and solve for t. Since $(x_2, y_2) \in E$ and $(x_1, y_1) \in E$ are points in l corresponding to $t = 1$ and $t = 0$ respectively, $t = 1, 0$ are roots of (3.3). Cancelling out groups of terms of the form $f(x_1, y_1) = 0$ and $f(x_2, y_2) = 0$, and dividing (3.3) by $t(t-1)$, we are left with

$$t(x_2 - x_1)^3 + (x_2 - x_1)^3 + (y_2 - y_1)^2 + 3x_1(x_2 - x_1)^2 = 0$$

So when $x_1 \neq x_2$,

$$t = -\frac{(x_2 - x_1)^3 - (y_2 - y_1)^2 + 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}$$

from which we calculate the third point of $E \cap l$, (x_0, y_0) to be

$$\begin{aligned} x_0 &= x_1 - \frac{(x_2 - x_1)^3 - (y_2 - y_1)^2 + 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}(x_2 - x_1) \\ &= -x_2 - x_1 + \frac{(y_2 - y_1)^2}{(x_2 - x_1)^2} \\ y_0 &= y_1 - \frac{(x_2 - x_1)^3 - (y_2 - y_1)^2 + 3x_1(x_2 - x_1)^2}{(x_2 - x_1)^3}(y_2 - y_1) \\ &= 2y_1 - y_2 + \frac{(y_2 - y_1)^3}{(x_2 - x_1)^3} - 3x_1\frac{(y_2 - y_1)}{(x_2 - x_1)} \end{aligned} \tag{3.5}$$

In the case $x_2 = x_1$, l is given by $x = x_1$ and the projective closure $L \subset \mathbb{P}^2(K)$ is given by $X - x_1 Z = 0$. Substituting in (2.3), we have

$$x_1^3 Z^3 - a x_1 Z^3 + b Z^3 - Y^2 Z = y_1^2 Z^3 - Y^2 Z = Z(y_1 Z - Y)(y_1 Z + Y)$$

which means that in $\mathbb{P}^2(K)$, $E \cap L = \{[0:1:0], [x_1 : y_1 : 1], [x_1 : -y_1 : 1]\}$. Looking at E as an affine curve along with $\mathcal{O}$, the points of intersection of E and l are (x_1, y_1), $(x_2, y_2) = (x_1, -y_1)$ and $\mathcal{O}$.

3.4.3 Lines intesecting $\mathcal{O}$

Let L be the line $Z = 0$ in $\mathbb{P}^2(K)$. To find $L \cap E$, we substitute $Z = 0$ into (2.3) to get $X^3 = 0$, of which $X = 0$ is a triple root. Therefore L touches E at $\mathcal{O}$ with multiplicity 3.

In addition, as we noted in Lemma 3.6, $[0:1:0]$ belongs to a line L iff L is given by

$$\alpha X + \gamma Z = 0$$

If L is any line other than the line at infinity, then $\alpha \neq 0$, and we can fix $\alpha = 1$. Then we substitute $X = -\frac{\gamma}{\alpha} Z$ in (2.3) and setting $Z = 0$ gives $\mathcal{O}$, and $Z = 1$ gives $Y^2 = f(\gamma, 0)$, which, if it has one root, has two (counting multiplicities). If $[x_0 : y_0 : 1]$ is one of these roots, the other is $[x_0 : -y_0 : 1]$. Another consequence it that the "tangent at infinity", the only line in $\mathbb{P}^2(K)$ that intersects $\mathcal{O}$ with multiplicity greater than one, is $Z = 0$.

Observe that all that we have established so far is a consequence of the fact that if a cubic equation in one variable over any field has two roots, then it has all three. This guarantees that if a line intersects E twice, it

also intersects E a third time, either in $\mathbb{A}^2(K)$ or at $\mathcal{O}$. We have also seen that a line containing $\mathcal{O}$ intersects $E \setminus \mathcal{O}$ only at pairs of points (x_1, y_1) and $(x_1, -y_1)$.

3.5 The Group Law

The remarks at the end of the previous section point to there being some structure to the set of points on an elliptic curve E. Given two points $P, Q \in E$, R, the third point of intersection of E with the line joining P and Q (or with the tangent at P if $P = Q$), is a well-defined binary operation on E. We shall denote it by $*$, or say that $P * Q = R$.

Definition 3.7. Let $P, Q \in E \subset \mathbb{P}^2(K)$, and let L_{PQ} be the line in $\mathbb{P}^2(K)$ containing P and Q (or the tangent line to E at P if $P = Q$). Let R be the third point of intersection (counting multiplicities) of E and L_{PQ}. Then the binary operation '$*$' is defined as

$$* : E \times E \longrightarrow E$$

$$* : (P, Q) \longmapsto R$$

We write $P * Q = R$ for $*(P, Q)$.

It is also clear from (3.4) and (3.5) that the coordinates of the third point belong to K, so E is closed with respect to $*$. Also, for a given choice of coordinates, $\mathcal{O}$ acts like a distinguished element of E. Altogether, it is not implausible that E could have a group structure.

If E can be made into a group, then $\mathcal{O}$ is the most likely candidate for the identity element. Also, by the commutative nature of the binary operation that we are motivated by, it is reasonable to assume that our group is abelian. Also, if $P = (x_1, y_1) \in E$, then $P' = (x_1, -y_1)$ is a natural candidate for the inverse element $-P$.

If we fix $\mathcal{O}$ to be the identity element, then $*$ cannot be the group operation, since

$$\mathcal{O} * P = P'$$

which is not equal to P in general. But if we define $P' = -P$, we have $\mathcal{O} * P = -P$, or $\mathcal{O} + P = P = -(\mathcal{O} * P)$.

Proposition 3.8. *If E is an elliptic curve, then $+ : E \times E \longrightarrow E$, such that $P + Q = -(P * Q)$ is a binary operation, with $P + Q = Q + P$ for all $P, Q \in E$.*

The above follows from the fact that $(P, Q) \mapsto P * Q$ and $P \mapsto -P$ are well defined binary (and respectively unary) operations, and E is closed with respect to both. Also $+$ commutes because $*$ does.

The definition $P + Q = -(P * Q)$ makes E into a group. All axioms other than associativity are straightforward.

Proposition 3.9. *Let $P \in E$. Then $\mathcal{O} + P = P + \mathcal{O} = P$.*

Proof. $\mathcal{O} + P = -(\mathcal{O} * P) = -(-P) = \mathcal{O} * -P = P$
$P + \mathcal{O} = \mathcal{O} + P = P$ by 3.8. □

Proposition 3.10. *Let $P \in E$. Define $-P = \mathcal{O} * P$. Then*
$P + (-P) = (-P) + P = \mathcal{O}$.

Proof. Note that $-\mathcal{O} = \mathcal{O} * \mathcal{O} = \mathcal{O}$.
$\therefore P + (-P) = -P + P = -(P * -P) = -\mathcal{O} = \mathcal{O}$. □

3.5.1 Associativity of Addition

Proving the associativity of + is a lot more difficult. We can derive addition formulae in terms of the coordinates of points of E and use these to verify associativity. But we shall give a geometric argument which will prove associativity in the case where the points involved are "in general position". First, we prove special cases of some important results about projective plane curves. For the rest of this section, we allow a projective plane cubic C to be the zero set of *any* homogeneous cubic polynomial in $K[X, Y, Z]$, ie., we do not insist on irreducibility. So C may be a product of 3 lines, or the product of a line and a conic. In such cases, we call the irreducible curves included in C the *components* of C.

Our first result is Lemma 3.6 for any cubic.

Lemma 3.11. *(Special case of Bezout's Theorem, I) Let C be a cubic and L a line in $\mathbb{P}^2(K)$. If C and L have no common components, then $\#(C \cap L) \leq 3$.*

Proof. Let $C = \mathcal{Z}(F)$, where $F \in K[X, Y, Z]$ is homogeneous of degree 3. Let $L = \mathcal{Z}(H)$, where $H = \alpha X + \beta Y + \gamma Z$. One of the coefficients of H is nonzero, since otherwise, we get $H(\mathbb{P}^2(K)) = 0$. Say $\gamma \neq 0$. Then we can substitute $Z = -\frac{\alpha X + \beta Y}{\gamma}$ in F to get a homogeneous degree 3 cubic in two variables. Say

$$F[X, Y, -\frac{\alpha X + \beta Y}{\gamma}] = a_{30}X^3 + a_{21}X^2Y + a_{12}XY^2 + a_{03}Y^3$$

Now if $a_{30} = 0$, substituting $Y = 0$ gives $[1 : 0 : 0]$ as a solution, and substituting $Y = 1$ gives a quadratic in X which gives at most two solutions. If $a_{30} \neq 0$, there are no solutions corresponding to $Y = 0$, and at most three solutions corresponding to $Y = 1$. □

A *conic* in $\mathbb{P}^2(K)$ is the zero set of a homogeneous polynomial of degree 2 (or quadratic form) in $K[X, Y, Z]$. Recall that if $Char(K) \neq 2$, then all quadratic forms can be expressed as symmetric K-bilinear forms $K^3 \longrightarrow K$. A conic is *nondegenerate* if the associated bilinear form can be expressed by an invertible matrix in $GL_3(K)$. We can also use projective transformations on K^3 to effect a change of basis, and hence a change of variables.

Let D be a nondegenerate conic, given by the zero set of the quadratic form

$$G(X, Y, Z) = aX^2 + bY^2 + cZ^2 + 2dXY + 2fXZ + 2eYZ.$$

We shall prove that, given at least one non-zero solution to $G(X, Y, Z) = 0$ there exists a basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ such that $G(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) = x_1x_3 - \alpha x_2^2$ for some $\alpha \in K \setminus 0$ and all $x_1, x_2, x_3 \in K$. Note that the symmetric bilinear form associated with G is

$$\varphi : K^3 \times K^3 \longrightarrow K,$$

$$\varphi : (\mathbf{v}_1, \mathbf{v}_2) \mapsto \mathbf{v}_1 M \mathbf{v}_2'$$

where the matrix M is given by

$$M = \begin{pmatrix} a & d & f \\ d & b & e \\ f & e & c \end{pmatrix}$$

Let $\mathbf{e}_1 \in K^3 \setminus \mathbf{0}$ such that $\varphi(\mathbf{e}_1, \mathbf{e}_1) = G(\mathbf{e}_1) = 0$. Since G is nondegenerate, there exists $\mathbf{e}_3 \in K^3$ such that $\varphi(\mathbf{e}_3, \mathbf{e}_3) = 0$ and $\varphi(\mathbf{e}_1, \mathbf{e}_3) \neq 0$. We can multiply $\mathbf{e}_3$ by a suitable constant to get $2\varphi(\mathbf{e}_1, \mathbf{e}_3) = 1$. Now the span $< \mathbf{e}_1, \mathbf{e}_3 >$ of $\mathbf{e}_1$ and $\mathbf{e}_3$ is a vector subspace of K^3 of dimension 2, and since M is invertible, so is $< \mathbf{e}_1 M, \mathbf{e}_3 M >$. So there exists $\mathbf{e}_2 \in K^3 \setminus \mathbf{0}$ such that $\varphi(\mathbf{e}_1, \mathbf{e}_2) = \varphi(\mathbf{e}_3, \mathbf{e}_2) = 0$. Let $< \mathbf{e}M >^{\perp} := \{\mathbf{v} \in K^3 | \mathbf{e}M\mathbf{v}' = 0\}$. Now, if $\varphi(\mathbf{e}_2, \mathbf{e}_2) = 0$, then $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in < \mathbf{e}_2 M >^{\perp}$. Since $< \mathbf{e}_2 M >^{\perp}$ has dimension 2, and $\mathbf{e}_2 \in < \mathbf{e}_1 M >^{\perp} \setminus < \mathbf{e}_1 >$, we have $\mathbf{e}_3 \in < \mathbf{e}_1, \mathbf{e}_2 > = < \mathbf{e}_1 M >^{\perp}$, which is a contradiction. Therefore, $\varphi(\mathbf{e}_2, \mathbf{e}_2) \neq 0$, and for

$$K^3 \cong < \mathbf{e}_2 M >^{\perp} \oplus < \mathbf{e}_2 > \cong < \mathbf{e}_1 > \oplus < \mathbf{e}_2 > \oplus < \mathbf{e}_3 >$$

Now, $\varphi(\mathbf{e}_1, \mathbf{e}_1) = \varphi(\mathbf{e}_3, \mathbf{e}_3) = \varphi(\mathbf{e}_1, \mathbf{e}_2) = \varphi(\mathbf{e}_3, \mathbf{e}_2) = 0$ and $\varphi(\mathbf{e}_1, \mathbf{e}_3) = \frac{1}{2}$. Let $\varphi(\mathbf{e}_2, \mathbf{e}_2) = -\alpha$. So

$$\begin{aligned} G(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) &= \varphi(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3, x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) \\ &= x_1^2\varphi(\mathbf{e}_1, \mathbf{e}_1) + x_2^2\varphi(\mathbf{e}_2, \mathbf{e}_2) + x_3^2\varphi(\mathbf{e}_3, \mathbf{e}_3) \\ &\quad + 2x_1x_2\varphi(\mathbf{e}_1, \mathbf{e}_2) + 2x_1x_3\varphi(\mathbf{e}_1, \mathbf{e}_3) + 2x_2x_3\varphi(\mathbf{e}_2, \mathbf{e}_3) \\ &= x_1x_3 - \alpha x_2^2 \end{aligned}$$

So any conic in the projective plane is equivalent to $G = XZ - \alpha Y^2$ for some $\alpha \in K \setminus 0$. Moreover, we can apply the transformation $X \mapsto \alpha X$, to get $G = XZ - Y^2$. We shall from now on assume that any conic D in the projective plane is the zero set of $G = XZ - Y^2$.

Lemma 3.12. *(Special case of Bezout's Theorem, II) Let C be a cubic and D a conic in $\mathbb{P}^2(K)$. If C and D have no common components, then $\#(C \cap D) \leq 6$.*

Proof. If D is degenerate, points of D lie either on a line or on the union of two lines. Since C and D cannot have common components, and since $\#(C \cap L) \leq 3$ for any line L, we have $\#(C \cap D) \leq 6$. If D is nondegerate, we can choose co-ordinates such that $D = XZ - Y^2$. Let $f_i(X, Z) \in K[X, Y]$ be homogeneous of degree i, with $0 \leq i \leq 3$, such that

$$C = f_0(X, Z)Y^3 + f_1(X, Z)Y^2 + f_2(X, Z)Y + f_3(X, Z).$$

Substituting $Y^2 = XZ$ and rearranging, we have

$$(XZf_0(X, Z) + f_2(X, Z))Y = -(XZf_1(X, Z) + f_3(X, Z)).$$

So $D(X, Y, Z) = 0$ and $C(X, Y, Z) = 0$ only if

$$XZ(XZf_0(X, Z) + f_2(X, Z))^2 - (XZf_1(X, Z) + f_3(X, Z))^2 = 0 \quad (3.6)$$

Now if $Z = 0$, then $D = 0$ gives $Y = 0, X = 1$. $C = 0$ gives $f_3(1, 0) = 0$. Since f_3 is homogeneous of degree 3, the last condition amounts to the coefficient of X^3 being 0. In this case, substituting $Z = 1$ in (3.6) gives at most a quintic in X, which has at most 5 solutions.

Substituting $Z = 1$ in (3.6) gives at most a sextic in X, which has at most 6 solutions. For each value of X such that $Xf_0(X, 1) + f_2(X, 1) \neq 0$, we have

$$Y = \frac{Xf_1(X, 1) + f_3(X, 1)}{XZf_0(X, Z) + f_2(X, Z)}.$$

which gives one value of Y corresponding to each value of X.

If $X = a$ is a solution of $Xf_0(X, 1) + f_2(X, 1) = 0$, then it is also a solution of $Xf_1(X, 1) + f_3(X, 1) = 0$. Therefore, if $Xf_0(X, 1) + f_2(X, 1) = (X - a_1)(X - a_2)$, then $Xf_1(X, 1) + f_3(X, 1) = (X - a_1)(X - a_2)p(X)$,

where $p(X)$ is at most linear in X. So for $Z = 1$, from (3.6) we have $(X - a_1)^2(X - a_2)^2(p(X)^2 - X) = 0$. Substituting $X = a_1, a_2$ in $D = 0$ gives at most 2 solutions each. So we again have a total of at most 6 points in the intersection. □

A similar argument goes through to show that two conics D_1 and D_2 with no common component intersect in at most four points in the projective plane. Moreover, any five points $P_1, \ldots, P_5 \in \mathbb{P}^2(K)$ lie on a conic. To see this, if three of the points lie on a line L, and if M is the line joining the other two, then $\{P_1, \ldots, P_5\} \subset LM$. So suppose no three points are collinear, and let $L_{12} = \overline{P_1P_2}$, $L_{34} = \overline{P_3P_4}$, $L_{13} = \overline{P_1P_3}$, and $L_{24} = \overline{P_2P_4}$. Then P_5 is not on any of the above lines, ie. $L_{12}(P_5)L_{34}(P_5)L_{13}(P_5)L_{24}(P_5) \neq 0$. So if

$$\gamma = -\frac{L_{13}(P_5)L_{24}(P_5)}{L_{12}(P_5)L_{34}(P_5)},$$

then $\{P_1, \ldots, P_5\} \subset \gamma L_{12}L_{34} + L_{13}L_{24}$. So $P_1, \ldots, P_5$ lie on a unique conic in $\mathbb{P}^2(K)$.

We can think of the set $\mathcal{C}$ of all homogeneous cubics in $K[X, Y, Z]$,(or cubic curves in $\mathbb{P}^2(K)$), together with the 0 polynomial, as a 10 dimensional vector subspace of $K[X, Y, Z]$ generated by the monomials of degree 3 in X, Y and Z. Since $\mathcal{Z}(F[X, Y, Z]) = \mathcal{Z}(\alpha F[X, Y, Z])$ for all $\alpha \in K$, we have $\mathcal{C} \cong \mathbb{P}^9(K)$. The set $\mathcal{C}(P_1)$ of all such curves passing through a given point $P_1 \in K^3 \setminus \mathbf{0}$, is a hyperplane of this projective space. If $P_2 \neq P_1$, then the set $\mathcal{C}(P_1, P_2)$ of cubics passing through both P_1 and P_2 is a hyperplane in $\mathcal{C}(P_1)$. Each further point which the set of cubics has to pass through imposes a further linear condition on the coefficients of the cubic. If the linear

conditions are independent, each extra point brings down the dimension of the subspace of cubics passing through these points by 1. The next result shows that if these points are added "in general position", we can add up to 8 points such that the corresponding linear conditions imposed are indeed independent.

Lemma 3.13. *Let $P_1, \ldots, P_8$ be eight points in the projective plane such that no four are collinear, and no seven lie on a conic. Then there is a cubic that contains $P_1, \ldots, P_7$ but not P_8.*

Proof. It is enough to show construct a cubic C which contains $P_1, \ldots, P_7$ but not P_8. C can be constructed as the product of a conic and a line, or the product of three lines depending on the configurations of the points.

Case 1. No three points in $\{P_1, \ldots, P_7\}$ are collinear.

The lines $\overline{P_4P_7}$, $\overline{P_5P_7}$, $\overline{P_6P_7}$ are distinct and intersect only at P_7. Hence at most one of these can contain P_8. Say $L = \overline{P_5P_7}$, and $L' = \overline{P_6P_7}$ do not contain P_8. Let K and K' respectively be the conics through $P_1, \ldots, P_5$ and $P_1, \ldots, P_4, P_6$. If both K and K' contain P_8, then they would intersect in five points $P_1, \ldots, P_4, P_8$, and we would have $K = K' \subset \{P_1, \ldots, P_6, P_8\}$. But this is impossible since seven points of the given eight cannot be contained in a conic, so one of K, K', say K cannot contain P_8. Then $KL' = 0$ contains $P_1, \ldots, P_7$.

Case 2. There is exactly one set of three collinear points in $\{P_1, \ldots, P_7\}$.

Assume the collinear points are $P_1, P_2, P_3 \in L$, where L is a line. Then no three points in $P_4, \ldots, P_7$ are collinear, and the lines $M = \overline{P_4P_5}$, $M' = \overline{P_4P_6}$, $N = \overline{P_6P_7}$, $N' = \overline{P_5P_7}$ are all distinct. The line pairs MN and $M'N'$ do not

both contain P_8, else either M' or N' would intersect one of M, N in two points. Assume MN does not contain P_8. Then $LMN = 0$ is a cubic which contains $P_1, \ldots, P_7$ but does not contain P_8.

Case 3. There are two distinct sets of three collinear points in $P_1, \ldots, P_7$.

Let L, M be lines such that $P_1, P_2, P_3 \in L$ and $P_4, P_5, P_6 \in M$. Then the lines $N = \overline{P_1 P_7}$ and $N' = \overline{P_2 P_7}$ are distinct, and cannot both contain P_8. Say $P_8 \notin N$. Then $LMN = 0$ is a cubic containing $P_1, \ldots, P_7$ but not P_8. These are all possible cases and the lemma is proved. □

It follows that if the points $P_1, \ldots, P_8$ are sufficiently general, then the set $\mathcal{C}(P_1, \ldots, P_8) \lneq \mathcal{C}(\{P_1, \ldots, P_8\} \setminus P_i)$ for all $1 \leq i \leq 8$.

Corollary. Let $P_1, \ldots, P_8$ be eight points in the projective plane such that no four are collinear, and no seven lie on a conic. Then $\mathcal{C}(P_1, \ldots, P_8) \cong \mathbb{P}^1(K)$.

Lemma 3.14. *Let C_1 and C_2 be any two projective plane cubics, with no component in common, that intersect in exactly 9 distinct points in $\mathbb{P}^2(K)$. Also suppose $P_1, \ldots, P_8$ are points in the intersection such that no four are collinear and no seven lie on a conic. Then any cubic $C_0 \in \mathcal{C}(P_1, \ldots, P_8)$ also passes through the ninth point P_9.*

Proof. Consider the cubic $C = \mathcal{Z}(F)$, where

$$\begin{aligned} F(X,Y,Z) &= a_{300}X^3 + a_{030}Y^3 + a_{003}Z^3 + a_{210}X^2Y + a_{201}X^2Z + a_{120}XY^2 \\ &\quad + a_{102}XZ^2 + a_{021}Y^2Z + a_{012}YZ^2 + a_{111}XYZ. \end{aligned} \tag{3.7}$$

If $\{P_1, \ldots, P_8\} \subset C$, then for each point $P_n = [x_n : y_n : z_n] \in \mathbb{P}^2(K)$, $F(P_n)$ is a linear equation in 10 variables a_{ijk}.

By Corollary to Lemma 3.13, $F(P_n) = 0,\ 1 \leq n \leq 8$ are linearly independent equations. Therefore, the set of all cubics passing through $\{P_1, \ldots, P_8\}$ is isomorphic the solution set of a system of 8 linearly independent equations in 10 variables. Then all cubic passing through $\{P_1, \ldots, P_8\}$ can be expressed as a linear combination of any two independent solutions of (3.7). Therefore, if $C_i = \mathcal{Z}(F_i),\ 0 \leq i \leq 2$, then $F_0 = \nu_1 F_1 + \nu_2 F_2$ for some constants ν_1, ν_2. Now since $F_1(P_9) = F_2(P_9) = 0$, we have $F_0(P_9) = 0$, ie. $P_9 \in C_0$. □

Proposition 3.15. *Let $A, B, C \in E$. Let $A+B = P$, $B+C = Q$, $P+C = R$, and $A + Q = R'$. If $\mathcal{S} = \{A, B, C, P, -P, Q, -Q, \mathcal{O}\}$ is a set of eight distinct points, and $-R, -R' \notin \mathcal{S}$, then $(A + B) + C = R = R' = A + (B + C)$.*

Proof. It is enough to prove that $-R = -R'$. Let L_1, L_2, L_3 be lines such that $A, B, -P \in L_1$, $\mathcal{O}, P, -P \in L_2$, and $P, C, -R \in L_3$. Let M_1, M_2, M_3 be lines such that $B, C, -Q \in M_1$, $\mathcal{O}, Q, -Q \in M_2$, and $A, Q, -R' \in M_3$. Since each of the above lines intersect E in three points each, and these three points are different for each line, Each of the lines $L_1, L_2, L_3, M_1, M_2, M_3$ are distinct. Define $C_1 = L_1 M_2 L_3$ and $C_2 = M_1 L_2 M_3$. Now

$$C_1 \cap E \supseteq \{A, B, C, \mathcal{O}, P, -P, Q, -Q, -R\}$$

$$C_1 \cap E \supseteq \{A, B, C, \mathcal{O}, P, -P, Q, -Q, -R'\}$$

$$C_1 \cap C_2 \supseteq \{A, B, C, \mathcal{O}, P, -P, Q, -Q\} = \mathcal{S}$$

Suppose $\mathcal{S}$ is a set of eight distinct elements.

Now, no four points in $\mathcal{S} \subset E$ are collinear. We claim that no seven points in $\mathcal{S}$ are on the same conic. To prove this, note that a line intersects

a conic in at most two points. So if $T \subset \mathcal{S}$ is a set of seven points lying on a conic, then not all three of $\mathcal{O}, P, -P \in L_2$ or $\mathcal{O}, Q, -Q$ are in T. Therefore, $\mathcal{O} \notin T$, or $T = \mathcal{S} \setminus \mathcal{O}$. But then A, B and $-P$ would be contained in $L_1 \cap T$. So we can apply Lemma 3.14.

But by Lemma 3.14, $-R \in C_2$, and $-R' \in C_1$. So

$$C_1 \cap C_2 \supseteq \mathcal{S} \cup \{-R, -R'\}$$

Also assume $-R, -R' \notin \mathcal{S}$. But since C_1, C_2 are products of three lines,

$$\#(C_1 \cap C_2) \leq \#(L_1 \cap C_2) + \#(M_2 \cap C_2) + \#(L_3 \cap C_2) \leq 3 + 3 + 3 = 9$$

The only possibility is $-R = -R'$, which gives $R = R'$. □

Now all that is left is to verify associativity where duplicates are involved, ie. when $\mathcal{S} = \{A, B, C, \mathcal{O}, P, -P, Q, -Q\}$ has less than eight elements, or when $-R, -R' \in \mathcal{S}$.

The above results can be summed up as follows.

Theorem 3.16. *Let E be an elliptic curve. Let $\mathcal{O}$ be the point at infinity on E, and $*$ be as in Definition 3.7. Let $+$ be the binary operation given by $P + Q = \mathcal{O} * (P * Q)$ for all $P, Q \in E$. Then $(E, +)$ is an abelian group.*

Proof. By Proposition 3.8, the operation $+$ is well defined and commutative on $E \times E$. By Proposition 3.9, $\mathcal{O}$ is the identity element. For all $P \in E$, $-P$ described in Proposition 3.10 is the inverse element. Associativity is established in Proposition 3.15 and the ensuing discussion. □

Chapter 4

The Elliptic Curve $E(\mathbb{F}_q)$

Let E be the elliptic curve given by (2.2) over a field k of characteristic $\neq 2, 3$. Let us recall the formulae we derived as coordinates of intersection in Section 3.4.

Theorem 4.1. *Let $P = [x_1 : y_1 : 1]$, $Q = [x_2 : y_2 : 1]$ be points on the elliptic curve E with $x_2 \neq x_1$. Let $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$. Then the coordinates of $P + Q = [x_0 : y_0 : 1]$ are*

$$\begin{aligned} x_0 &= \lambda^2 - x_2 - x_1 \\ y_0 &= y_2 - 2y_1 + \lambda^3 + 3x_1\lambda \\ &= (y_2 - \lambda x_2) - 2(y_1 - \lambda x_1) - \lambda x_0 \end{aligned}$$

Let $y_1 \neq 0$, and $\lambda_1 = \frac{3x_1^2 + a}{2y_1}$. Then the coordinates of $2P = [x_1' : y_1' : 1]$ are

$$\begin{aligned} x_1' &= \lambda_1^2 - 2x_1 \\ y_1' &= -\lambda_1^3 + 3x_1\lambda_1 - y_1 \\ &= -(y_1 - \lambda_1 x_1) - \lambda_1 x_1' \end{aligned}$$

Proof. Follows from (3.5) and (3.4). □

The first of the above formulae is called the addition formula while the second is called the duplication formula.

We can also answer the following question, which is in some sense an inverse of the duplication formula. When is a point P on an elliptic curve twice another point? Though not necessarily all points on an elliptic curve E are of the form $2Q$ for some $Q \in E$, we can prove that such a point exists if we extend the curve to contain points given by solutions to (2.2) over some field extension.

Theorem 4.2. *If P is a point on an elliptic curve $E(k)$, then there exists a point $P \in E(\bar{k})$ such that $P = 2Q$*

Proof. Let $P = (x_0, y_0)$. By the duplication formula, any point $Q = (x, y)$ such that $P = 2Q$ satisfies

$$\left(\frac{3x^2 + a}{2y}\right)^2 - 2x = x_0$$

We can substitute (2.2) and rewrite this as

$$\begin{aligned} (3x^2 + a)^2 - 4(x^3 + ax + b)(2x + x_0) &= 0 \\ \implies x^4 - 4x_0x^3 - 2ax^2 - (8b + 4ax_0)x + (a^2 - 4x_0b) &= 0 \end{aligned}$$

This is a quartic in x and has roots in some algebraic extension of k. If x_1 is a root, the y-coordinates are given by solutions to

$$(x_1^2 - a)^2 - 8bx_1 = 4x_0y^2$$

, which also exist is an extension field of k. □

4.1 Points of small order

Coordinates and number of the points of small orders in E are easy to calculate from Theorem 4.1 above.

If $(x_0, y_0) = P \in E$ is a point of order 2, then $P = -P$, and hence $y_0 = 0$. Therefore x_0 is a root of $x^3 + ax + b = 0$. Conversely, if $x_0^3 + ax_0 + b = 0$, then $(x_0, 0) \in E$ is a point of order 2. Also note that E has either 0, 1 or 3 points of order 2, depending on the number of roots of $x^3 + ax + b$.

If $(x_0, y_0) = P \in E$ is a point of order 3, then $2P = -P$. The tangent to E at P meets E only at P. Substituting in the duplication formula above gives

$$x_0 = \left(\frac{3x_0^2 + a}{2y_0}\right)^2 - 2x_0$$

which, on substituting $y_0^2 = x_0^3 + ax_0 + b$ simplifies to

$$3(x_0^2 + a)^2 - 4(a^2 - 3x_0 b) = 0 \tag{4.1}$$

The above quartic equation in x_0 has 0, 1, 2 or 4 solutions. When $char(\mathbb{F}) \neq 2$, each x-coordinate corresponds to 0 or 2 y-coordinates, since $y_0 \neq 0$. So the number of points of order 3 in E is 0, 2, 4, 6 or 8, and hence at most 8. But the points of order 3, along with the identity $\mathcal{O}$ forms the kernel of the homomorphism $P \mapsto 3P$, and hence a subgroup of the finite abelian group E, which we shall call E_3. Therefore E_3 is trivial, cyclic, or elementary abelian of rank 2, i.e, isomorphic to $<>$, C_3, or $C_3 \times C_3$. Therefore E has 0, 2, or 8 points of order 3.

If $(x_1, y_1) = Q \in E$ is a point of order 4, then $2Q$ is a point of order 2.

Let $2Q = P = (x_0, 0)$. Then the duplication formula gives

$$\begin{aligned} x_0 &= \left(\frac{3x_1^2 + a}{2y_1}\right)^2 - 2x_1 \\ 0 &= \frac{3x_1^2 + a}{2y_1}(x_0 - x_1) + y_1 \end{aligned}$$

We can substitute $y_1^2 = x_1^3 + ax_1 + b$, then expand and simplify the second equation as

$$\begin{aligned} 0 &= (3x_1^2 + a)(x_1 - x_0) - 2(x_1^3 + ax_1 + b) \\ &= x_1^3 - 3x_0x_1^2 - ax_1 - ax_0 - 2b \\ &= (x_1 - x_0)^3 - 3x_0^2(x_1 - x_0) - a(x_1 - x_0) - 2(x_0^3 + ax_0 + b) \\ &= (x_1 - x_0)^3 - (3x_0^2 + a)(x_1 - x_0) \qquad (4.2) \end{aligned}$$

since $x_0^3 + ax_0 + b = 0$. So we have $x_1 = x_0$, and $x_1 = x_0 \pm \sqrt{(3x_0^2 + a)}$ if $3x_0^2 + a$ is a square in K. But the solution $x_1 = x_0$ is redundant. So for each of the three possible points P of order 2, there exists at most 4 points Q such that $2Q = P$. So we have a total of at most 12 points of order 4 in E. This tells us that the subgroup formed by elements of order dividing 4 is a subgroup of $C_4 \times C_4$.

4.2 The curve E_{q^2}

Let E_q be the curve given by the equation $x^3 + ax + b = y^2$, where $a, b \in \mathbb{F}_q$. We assume that $f(x) = x^3 + ax + b$ does not have a multiple root in $\bar{\mathbb{F}}_q$, and that if $char(\mathbb{F}_q) = p$, then $p \neq 2, 3$.

Define E_{q^2} to be the curve given by the same equation as above, but the x and y coordinates take values from $\mathbb{F}_{q^2}$. Our assumptions on $f(x) \in \mathbb{F}_{q^2}[x]$ and p still hold.

4.2.1 The Frobenius Map

The Frobenius automorphism ϕ on $\mathbb{F}_{q^2}$, defined as

$$\begin{aligned} \phi: \quad \mathbb{F}_{q^2} &\rightarrow \mathbb{F}_{q^2} \\ \phi: \quad x &\mapsto x^q \end{aligned}$$

which restricts to the identity map on $\mathbb{F}_q$, induces a map

$$\begin{aligned} \Phi: \quad E &\rightarrow E \\ \Phi: \quad (x,y) &\mapsto (x^q, y^q) \end{aligned}$$

Since ϕ is a field automorphism, and formulae for sums and inverses of elements of E are all rational expressions, Φ is a group homorphism from E to E and also from E_q to E_q for all $E_q < E$

Note that $\Phi^2 = \mathbb{I}$, the identity automorphism on E_{q^2}. Also, restricting Φ to E_q gives the identity map. More specifically, $\Phi(P) = P$ if and ony if $P \in E_q \quad \forall\ P \in E_{q^2}$.

4.2.2 The Sturucture of $2E_{q^2}$

Consider $2E_{q^2} := \{2P \mid P \in E_{q^2}\}$, the image of E_{q^2} under the group homomorphism

$$\begin{aligned} \varphi: \quad E_{q^2} &\rightarrow E_{q^2} \\ \varphi: \quad P &\mapsto 2P \end{aligned}$$

which has $ker(\varphi) = \{Q \in E_{q^2} \mid 2Q = \mathcal{O}\}$, ie., $\mathcal{O}$ and the points of order 2 in E_{q^2}. Let $\#\{Q \in E_q \mid 2Q = \mathcal{O}\} = d_1$ and $\#\{Q \in E_{q^2} \mid 2Q = \mathcal{O}\} = d_2$. $d_1, d_2 \in \{1, 2, 4\}$, $d_1 \leq d_2$. In fact, we can say more.

If f is irreducible over $\mathbb{F}_q$, then f is irreducible over $\mathbb{F}_{q^2}$. Also, if f has one root in $\mathbb{F}_q$, then f has three roots in $\mathbb{F}_{q^2}$.

$$\begin{aligned} d_1 = 1 \quad &\implies \quad d_2 = 1, \\ d_1 = 2 \quad &\implies \quad d_2 = 4 \\ d_1 = 4 \quad &\implies \quad d_2 = 4 \end{aligned}$$

Observe that $\forall\ P \in E_{q^2}, \quad 2P = P + \Phi(P) + P - \Phi(P)$.

Let $\mathcal{R} := \{P + \Phi(P) \mid P \in E_{q^2}\}$ and $\mathcal{S}_1 := \{P - \Phi(P) \mid P \in E_{q^2}\}$. Then

$$2E_{q^2} \subseteq \mathcal{R} + \mathcal{S}$$

Also

$$\begin{gathered} \Phi(P + \Phi(P)) = \Phi(P) + \Phi^2(P) = P + \Phi(P) \\ \implies \forall\ P \in E_{q^2}, \quad P + \Phi(P) \in E_q \\ \text{and} \quad \forall\ P \in E_q \quad P + \Phi(P) = 2P \end{gathered}$$

Also, since E_q is an abelian group, $P+\Phi(P)+Q+\Phi(Q) = P+Q+\Phi(P+Q)$, so $\mathcal{R}$ is additively closed, and we have $E_q \geq \mathcal{R} \geq 2E_q$ and $|\mathcal{R}| = |E_q|\frac{d_0}{d_1}$ for some $d_0 \mid d_1$.

Let $\mathcal{S}' = \{Q \in E_{q^2} | \Phi(Q) = -Q\}$.

$$\forall P \in E_{q^2},\ \Phi(P - \Phi(P)) = \Phi(P) - \Phi^2(P) = -(P - \Phi(P))$$

So we have $\mathcal{S} \subseteq \mathcal{S}'$. Also, $Q \in \mathcal{S}' \iff (x^q, y^q) = (x, -y)$, which implies that $x \in \mathbb{F}_q$. Also, if α is a solution to $y^q + y = 0$, then all other solutions

are given by αz, where $z \in \mathbb{F}_q$. Since $\alpha^q = -\alpha$, we have $(\alpha^2)^q = \alpha^2$, ie. $\alpha^2 \in \mathbb{F}_q$. Thus $\alpha \in \mathbb{F}_{q^2}$, and the set of non-zero y-coordianates of elements of $\mathcal{S}'$ is the set of all square roots of values of $x^3 + ax + b$ which are quadratic non-residues in $\mathbb{F}_q$. Therefore, $\exists\ \alpha \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$, with $\alpha^2 \in \mathbb{F}_q$ such that $\forall\ Q \in \mathcal{S}',\ Q = (x, \alpha y),\ x, y \in \mathbb{F}_q$.

Now $Q \in \mathcal{S}' \implies \Phi(Q) = -Q$, and we have $Q - \Phi(Q) = Q - (-Q) = 2Q$. Thus $Q \in \mathcal{S}' \implies 2Q \in \mathcal{S}$ and we have $2\mathcal{S}' := \varphi(\mathcal{S}') \subset \mathcal{S}$. Also, it is clear from the addition and duplication formulas that $\mathcal{S}'$ is closed under elliptic curve addition and taking inverses, therefore $\mathcal{S}' < E_{q^2}$, and hence $2\mathcal{S}' < E_{q^2}$.

Note that $P - \Phi(P) = \mathcal{O} \iff P \in E_q$. Also, $(\mathbb{I} - \Phi)$ is a group homomorphism from E_{q^2} to itself, since $(\mathbb{I} - \Phi)(P + Q) = P + Q - \Phi(P + Q) = P - \Phi(P) + Q - \Phi(Q)$. The image of this homomorphism is $\mathcal{S}$. So we have $\mathcal{S} < E_{q^2}$.

Combining the two arguments above, we have $2\mathcal{S}' \leq \mathcal{S} \leq \mathcal{S}'$. Recall that $|\mathcal{S}'| = d_1 |2\mathcal{S}'|$. So we have $|\mathcal{S}| = |\mathcal{S}'| \frac{d_1'}{d_1}$, for some $d_1' \mid d_1$.

We can now count the number of points in E_{q^2} given the number of points in E_q.

Theorem 4.3. *Let* $|E_q| = q + 1 + \pi(f)$. *Then* $|E_{q^2}| = (q+1)^2 - \pi(f)^2$

Proof. First, observe that

$$\mathcal{S}' = \{\mathcal{O}\} \cup \{x \in \mathbb{F}_q | x^3 + ax + b = 0\} \cup \{x \in \mathbb{F}_q | x^3 + ax + b = \alpha^2 y^2\}$$

where α is as above. So

$$|\mathcal{S}'| = d_1 + 2\#\{x \in \mathbb{F}_q | x^3 + ax + b = \alpha^2 y^2,\ \alpha \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q,\ \alpha^2 \in \mathbb{F}_q\}$$

Similarly we have

$$|E_q| = d_1 + 2\#\{x \in \mathbb{F}_q | x^3 + ax + b = y^2,\ y \in \mathbb{F}_q^*\}$$

Therefore,

$$\begin{aligned} |\mathcal{S}'| + |E_q| &= 2d_1 + 2\#\{x \in \mathbb{F}_q | x^3 + ax + b \neq 0\} \\ &= 2 + 2\#\{x \in \mathbb{F}_q\} \\ &= 2(q+1) \end{aligned}$$

So if $|E_q| = q + 1 + \pi(f)$, then $|\mathcal{S}'| = q + 1 - \pi(f)$ Also, we have

$$2E_{q^2} \subseteq E_q + \mathcal{S}' \subseteq E_{q^2}, \tag{4.3}$$

since $2E_{q^2} \subseteq \mathcal{R} + \mathcal{S}$. We have the following three cases.

Case 1: x^3+ax+b has no roots in $\mathbb{F}_q$, ie. $d_1 = 1$. Then $E_{q^2} = 2E_{q^2} = \mathcal{R}+\mathcal{S}$. But $d_0 = d_1 = d_1' = 1$, so $\mathcal{R} = E_q$ and $\mathcal{S} = \mathcal{S}'$, $E_q \cap \mathcal{S}' = \{\mathcal{O}\}$. So we have $|E_{q^2}| = |E_q||\mathcal{S}'|$ and we are done.

Case 2: $x^3 + ax + b$ has three roots in $\mathbb{F}_q$, ie, $d_1 = d_2 = 4$. Here we have $|2E_{q^2}| = \frac{1}{4}|E_{q^2}|$, and $|E_q \cap \mathcal{S}'| = 4$. So it is enough to show that $2E_{q^2} = E_q + \mathcal{S}'$.

We know that $2E_{q^2} \subseteq E_q + \mathcal{S}'$. To show that $2E_{q^2} \supseteq E_q + \mathcal{S}'$ it is enough to show that $2E_{q^2} \supseteq E_q$ and $2E_{q^2} \supseteq \mathcal{S}'$.

Let $P \in E_q$. We need to show that $P = 2Q$ for some $Q \in E_{q^2}$. Now by Theorem 4.2, if $\bar{E} = E/\bar{\mathbb{F}}_q$, $\exists Q \in \bar{E}$ which satisfies $2Q = P$. Now, $2Q = P = \Phi(P) = \Phi(2Q) = 2\Phi(Q)$, which implies $2(Q - \Phi(Q)) = \mathcal{O}$. Now if $Q \notin E_q$, $\Phi(Q) \neq Q$, so $Q - \Phi(Q)$ has order 2, ie $\Phi(Q) = Q + T$, where T is an involution in E_q. But

$$\Phi^2(Q) = \Phi(Q + T) = \Phi(Q) + \Phi(T) = (Q + T) + T = Q$$

Therefore $Q \in E_{q^2}$, which is what we wanted. The exact same argument shows that $\mathcal{S}' \subseteq 2E_{q^2}$.

The remaining case is slightly more complicated.

Case 3: x^3+ax+b has one root in $\mathbb{F}_q$, ie. $d_1 = 2$. In this case the irreducible quadratic factor of $f(x)$ splits in $\mathbb{F}_{q^2}$. Let T_0 be the point of order 2 given by the root of $f(x)$ in $\mathbb{F}_q$, and let the two involutions in the extended curve be T_1 and T_2 . So $|\mathcal{S}'| \cap |E_q| = \#\{\mathcal{O}, T_0\} = 2$, which gives $|E_q + \mathcal{S}'| = \frac{|E_q||\mathcal{S}'|}{2}$. Now as $|2E_{q^2}| = \frac{1}{4}|E_{q^2}|$, if we show that both inclusions in (4.3) are proper, we will have

$$|E_q + \mathcal{S}'| = \frac{1}{2}|E_{q^2}|$$

which will give the desired result.

To prove that $2E_{q^2} \subsetneq E_q + \mathcal{S}'$, it is enough to show that for some $P \in E_q$, we have $P \neq 2Q \ \forall Q \in E_{q^2}$. Let P be any element of E_q such that $P \neq 2Q \ \forall Q \in E_q$, ie. P is not twice an element of E_q. (Such a P exists since E_q is finite of even order.) If $P = 2Q$ for some $Q \in E_{q^2}$, as in the previous case, we would have $2(Q - \Phi(Q)) = \mathcal{O}$. Since $Q - \Phi(Q) \in \mathcal{S}'$, we have $Q - \Phi(Q)$ is the involution E_q ie $Q - \Phi(Q) = T_0$. But $2(Q + T_1)$ is also equal to P. So repeating the above argument for $Q + T_1$, we have

$$\begin{aligned} \Phi(Q+T_1) &= Q + T_1 + T_0 \quad \text{and also} \\ \Phi(Q+T_1) &= \Phi(Q) + \Phi(T_1) \\ &= Q + T_0 + \Phi(T_1) \\ \implies \quad \Phi(T_1) &= T_1 \end{aligned}$$

which contradicts $T_1 \notin E_q$. So we cannot have $P = 2Q$ for $P \in E_q \setminus 2E_q$ and

$Q \in E_{q^2}$.

Now we need to show that $E_q + \mathcal{S}' \subsetneq E_{q^2}$. We know that $T_1, T_2 \in E_{q^2} \setminus (E_q \cup \mathcal{S}'$. It is enough to show that one of T_1, T_2 can not be written as $U + V$ where $U \in E_q$ and $V \in \mathcal{S}'$. Suppose $T_1 = U_1 + V_1$ and $T_2 = U_2 + V_2$, where $U_i \in E_q$ and $V_i \in \mathcal{S}'$ for $i = 1, 2$.

Now $\Phi(T_1) = T_2$ and also $\Phi(T_1) = \Phi(U_1) + \Phi(V_1) = U_1 - V_1$. Therefore

$$T_0 = T_1 + \Phi(T_1) = 2U_1 \text{and} T_0 = T_1 - \Phi(T_1) = 2V_1$$

Similarly, since $\Phi(T_2) = T_1$, and $\Phi(T_2) = \Phi(U_2) + \Phi(V_2) = U_2 - V_2$, we have

$$T_0 = T_2 + \Phi(T_2) = 2U_2 \text{and} T_0 = T_2 - \Phi(T_2) = 2V_2$$

Now since $U_1, U_2 \in E_q$, we have $U_1 + U_2 \in E_q$. But $2(U_1 + U_2) = 2T_0 = \mathcal{O}$. Therefore $U_1 + U_2 = T_0$, which is also equal to $2U_1$. So $U_1 = U_2$. Also, since $V_1, V_2 \in \mathcal{S}'$, we have $V_1 + V_2 \in \mathcal{S}'$, but as $2(V_1 + V_2) = \mathcal{O}$, we have $V_1 + V_2 = T_0 = 2V_1$, ie., $V_1 = V_2$. But $U_1 + V_2 = T_1 \neq T_2 = U_2 + V_2$, which is a contradiction. So at least one of $T_1, T_2 \notin E_{q^2}$.

This completes our proof.

□

Bibliography

[AM] M.F. Atiyah and I.G. Macdonald, Introduction to commutative algebra, Addison-Wesley, 1969.

[BIX] R. Bix, Conics and cubics: a concrete introduction to algebraic curves, Undergraduate texts in mathematics, Undergraduate texts in mathematics, Springer, 1998.

[ST] J.H. Silverman and J.T. Tate, Rational points on elliptic curves, Undergraduate texts in mathematics, Springer, 1992.

[REI] M. Reid Undergraduate algebraic geometry, Volume 12 of London Mathematical Society student texts, London Mathematical Society, Cambridge University Press, 1988.

[WAS] L.C. Washington, Elliptic curves: number theory and cryptography, CRC Press series on discrete mathematics and its applications, CRC Press, 2003.

[SHA] I.R. Shafarevich, Basic algebraic geometry, Volume 1 Springer Study Edition Series, Springer, 1994.